

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 9, 2002, 00:09:23 ; Search time 8498.8 Seconds  
(without alignments)  
29.081 Million cell updates/sec

Title: US-09-851-670-14

Perfect score: 23  
Sequence: 1 gagaaacacccgcctctcgcgcaaa 23

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 111874

Minimum DB seq length: 0  
Maximum DB seq length: 60

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estipl:\*  
6: em\_estiba:\*  
7: em\_estro:\*  
8: em\_estrov:\*  
9: em\_hlc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15.8	68.7	52	10	AM693240
2	14.6	63.5	50	10	AU107235
3	14	60.9	40	10	AU1859778
4	13.6	59.1	40	13	A2463268
5	13.4	58.3	34	13	AQ025206
6	13.4	58.3	49	13	A2514447
7	13.2	57.4	60	13	A2494986
8	13	56.5	25	13	A2303786
9	13	55.5	59	10	AA098077
10	12.8	55.7	55	10	AA894695
11	12.6	54.8	26	13	A2342914
12	12.6	54.8	26	13	TA194F010

c 13	12.6	54.8	35	13	A2845779	A2845779
c 14	12.6	54.8	50	10	AU106768	AU106768
c 15	12.4	53.9	32	13	A2462085	A2462085
c 16	12.4	53.9	45	13	A2635879	A2635879
c 17	12.4	53.9	46	10	A1811483	A1811483
c 18	12.4	53.9	49	10	AW432778	AW432778
c 19	12.4	53.9	57	10	AW424126	AW424126
c 20	12.4	53.9	57	13	AQ024986	AQ024986
c 21	12.4	53.9	58	10	A1499235	A1499235
c 22	12.2	53.0	21	13	A2581103	A2581103
c 23	12.2	53.0	21	13	A2857747	A2857747
c 24	12.2	53.0	36	13	BH011404	BH011404
c 25	12.2	53.0	40	13	TA2538010	TA2538010
c 26	12.2	53.0	49	10	A1792794	A1792794
c 27	12.2	53.0	49	10	AA429584	AA429584
c 28	12.2	53.0	50	10	AU103878	AU103878
c 29	12.2	53.0	50	10	AU103887	AU103887
c 30	12.2	53.0	50	10	AU106770	AU106770
c 31	12.2	53.0	52	11	BF643352	BF643352
c 32	12.2	53.0	58	13	A2920130	A2920130
c 33	12	52.2	30	13	AQ025600	AQ025600
c 34	12	52.2	41	11	H93873	H93873
c 35	12	52.2	43	10	A1245483	A1245483
c 36	12	52.2	46	10	A1744926	A1744926
c 37	12	52.2	46	13	A2810625	A2810625
c 38	12	52.2	49	10	AA663894	AA663894
c 39	12	52.2	50	13	AQ025192	AQ025192
c 40	12	52.2	51	13	AQ025045	AQ025045
c 41	12	52.2	51	13	A2575776	A2575776
c 42	12	52.2	52	10	BE321098	BE321098
c 43	12	52.2	52	11	BF631938	BF631938
c 44	11.8	51.3	26	13	TA65F020	TA65F020
c 45	11.8	51.3	31	10	A1383811	A1383811

#### ALIGNMENTS

RESULT 1  
AM693240 52 bp mRNA EST 20-DEC-2000  
LOCUS NF062A03ST1F1000 Developing stem Medicago truncatula cDNA clone  
DEFINITION NF062A03ST 5', mRNA sequence.

ACCESSION AM693240.2 GI:11933549

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE  
AUTHORS He X.-Z., Shadle G., Scott A.D., Harris A.R., Gonzales R.A., Bell C.J., Flores H.R., Imman J.T., Weller J.W., May G.D. and Dixon R.A.

TITLE  
JOURNAL  
COMMENT  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula stem library  
Unpublished (2000)  
On Apr 14, 2000 this sequence version replaced gi:7567976.  
Contact: Dixon RA

Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7302  
Fax: 580 221 7380

Email: radixon@noble.org  
Insert Length: 752 Std Error: 0.00  
Plate: 062 Row: A Column: 03  
Seq primer: TCACACAGCAACACCTGTGAC.

FEATURES  
source  
1..52  
/organism="Medicago truncatula"

/db\_xref="taxon:3880"  
/clone="NF062A03ST"  
/clone.lib="Developing stem"  
/tissue\_type="stem"  
/dev\_stage="Pooled developmental"  
/note="Vector: Lambda Zap; Contains a mixture of  
intermodal stem segments"

BASE COUNT 12 a 22 c 2 g 16 t  
ORIGIN

Query Match 68.7%; Score 15.8; DB 10; Length 52;  
Best Local Similarity 89.5%; Pred. No. 4.5e+03;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 aagaccgcctctcgcaa 22  
||||| ||||| |||  
DB 9 AACACCCACTCTCTCA 27

RESULT 2  
LOCUS AU107235 50 bp mRNA EST 05-APR-2001  
DEFINITION AU107235 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
LN607514, mRNA sequence.  
ACCESSION AU107235  
VERSION AU107235.1 GI:13556756  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo  
K., Suyama,A. and Sugano,S.  
Fine Structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
Unpublished (2001)  
JOURNAL Contact: Yutaka Suzuki  
COMMENT Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES  
source  
1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="LN607514"  
/clone.lib="Sugano Homo sapiens cDNA library"

BASE COUNT 16 a 15 c 9 g 10 t  
ORIGIN

Query Match 63.5%; Score 14.6; DB 10; Length 50;  
Best Local Similarity 81.0%; Pred. No. 1.4e+04;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 agaacaccgcctctcgcaa 22  
||||| ||||| |||  
DB 22 AGAACACCTCTCTCTCA 42

RESULT 3  
LOCUS AI859778 40 bp mRNA EST 07-MAR-2000  
DEFINITION AI859778 mm21e09.x1 NCI CGAP, U44 Homo sapiens cDNA clone IMAGE:243616 3'  
similar to TR:Q08380 Q08380 MAC-2 BINDING PROTEIN PRECURSOR. ;,  
mRNA sequence.  
ACCESSION AI859778

VERSION AI859778.1 GI:5513479  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA library Preparation: Life Technologies, Inc.  
cDNA library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

FEATURES  
source  
1..40  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:243616"  
/clone.lib="NCI CGAP, U44"  
/tissue\_type="serous papillary carcinoma, high grade, 2  
pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: PCMV-SPORT6; Site:1; Salt;  
Site:2; NotI: Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.48 kb. Life Technologies catalog #:  
11542-016"

BASE COUNT 13 a 13 c 9 g 5 t  
ORIGIN

Query Match 60.9%; Score 14; DB 10; Length 40;  
Best Local Similarity 77.3%; Pred. No. 2.6e+04;  
Matches 17; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 agaacaccgcctctcgcaa 23  
||||| ||||| |||  
DB 7 AGAACACCTCTCTCTCA 28

RESULT 4  
LOCUS AZ463268 40 bp DNA GSS 04-OCT-2000  
DEFINITION IM0272B01F Mouse 10kb plasmid UGCCIM library Mus musculus genomic  
clone UGCCIM0272B01 F, DNA sequence.  
ACCESSION AZ463268  
VERSION AZ463268.1 GI:10621393  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 40)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.  
and Wright,D. Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss

University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0272 row: B column: 01  
 Seq primer: CGTGTAAACGACGCCACGT  
 Class: plasmid ends  
 High quality sequence stop: 40.

## FEATURES

Source

1. 40  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUCG1M0272B01"  
 /clone\_lib="Mouse 10kb plasmid UUCG1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (9114732114[gblAF129072.1]), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT  
 ORIGIN

5 a 20 c 2 g 13 t

## Query Match

Best Local Similarity 59.1%; Score 13.6; DB 13; Length 40;  
 Matches 16; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gagacaccgcgtctctcgc 20  
 ||| ||||| ||||| |||  
 Db 2 GAAACACCCCTCTCTCTC 21

RESULT 5  
 A0025206/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

A0025206 34 bp DNA GSS 23-AUG-2000  
 EP131249 Drosophila melanogaster EP line Drosophila melanogaster  
 genomic Sequence recovered from 5' end of P element, DNA sequence.  
 A0025206  
 A0025206.1 GI:3265558  
 GSS.  
 fruit fly,  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 34)  
 Liao, G.-C., Rehm, E. J. and Rubin, G. M.  
 Insertion site preferences of the P transposable element in  
 Drosophila melanogaster  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3347-3351 (2000)  
 20202638  
 Contact: Gerald Rubin

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT

Berkeley Drosophila Genome Project  
 University of California, Berkeley  
 LSA Building, Berkeley, CA 94720-3200, USA  
 Fax: 5106433947  
 Email: gerry@fruitfly.berkeley.edu  
 Sequence recovery method was inverse PCR.

Sequence orientation is forward strand relative to 5' end of P element

The P element insertion position is base 27 in the 34 bases. This insertion position refers to the first base of the 8 base target recognition sequence.

Class: transposon-tagged.

## FEATURES

Source

Location/Qualifiers  
 1. 34  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone\_lib="Drosophila melanogaster EP line"  
 /note="Inverse PCR was performed on Drosophila melanogaster strains each of which contains a single EP transposable element insertion. (The generation of these insertion strains is described in Roth P, Szabo K, Bailey A, Laverly T, Rehm J, Rubin GM, Weigmann K, Milan M, Benes V, Ansoorge W, Cohen SM. 1998. Systematic gain-of-function genetics in Drosophila. Development 6:1049-1057.) The resultant fragment for each strain was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://fruitfly.berkeley.edu/P-disrupt/inverse\_pcr.html."

BASE COUNT  
 ORIGIN

4 a 5 c 17 g 8 t

## Query Match

Best Local Similarity 58.3%; Score 13.4; DB 13; Length 34;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 aacaccgcgtctctc 18  
 ||| ||||| ||||| |||  
 Db 28 ACCACCCGCTCTCTC 14

RESULT 6  
 A2514447  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

A2514447 49 bp DNA GSS 05-OCT-2000  
 IM0361M14F Mouse 10kb plasmid UUCG1M library Mus musculus genomic  
 clone UUCG1M0361M14 F, DNA sequence.  
 A2514447  
 A2514447.1 GI:10695859  
 GSS.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurgnathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 49)  
 Dunn, D., Aoyagi, A., Barber, M., Baecorn, T., Duval, B., Hamil, C.,  
 Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
 M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausern, A.  
 and Wright, D., Weiss, R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0361 row: M column: 14

TITLE  
 JOURNAL  
 MEDLINE  
 COMMENT





/clone="IMAGE:1493385"  
 /clone\_lib="NCI\_CGAP\_kid3"  
 /lab\_host="DH10B"  
 /note="Organ: kidney; Vector: pRTT3D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTT3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 9 a 6 c 30 g 9 t 1 others  
 ORIGIN

Query Match 55.7%; Score 12.8; DB 10; Length 55;  
 Best Local Similarity 87.5%; Pred. No. 8.6e+04;  
 Matches 14; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 acaaccgcctctcgc 20  
 ||||| |||||  
 Db 22 ACACCCCTCTCCTCGC 7

RESULT 11  
 A2342914 26 bp DNA GSS 29-SEP-2000  
 LOCUS A2342914  
 DEFINITION clone UUGC1M0076C22 f, DNA sequence.  
 ACCESSION A2342914  
 VERSION A2342914.1 GI:10420628  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 26)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausen,A., and Wright,D., Weiss,R.

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunne@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00  
 Plate: 0076 row: C column: 22  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 26.  
 Location/Qualifiers  
 1..26  
 /organism="Mus musculus"  
 /strain="C57BL/6j"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0076C22"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6j (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA

was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g1473211419b/AP129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 3 a 11 c 1 g 11 t  
 ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 26;  
 Best Local Similarity 78.9%; Pred. No. 9.6e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 agaacaccgcctctcgc 20  
 ||||| | | ||||| |  
 Db 1 AGAAGCTCTCTCTCTCTC 19

RESULT 12  
 TA194F010/c 26 bp DNA GSS 13-DEC-2000  
 LOCUS TA194F010/c  
 DEFINITION T. Brucei sheared genomic DNA clone 194f01, reverse sequence, genomic survey sequence.  
 ACCESSION AL477302  
 VERSION AL477302.1 GI:11841328  
 KEYWORDS GSS.  
 SOURCE Trypanosoma brucei.  
 ORGANISM Trypanosoma brucei  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

REFERENCE 1 (bases 1 to 26)  
 Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.  
 Direct Submission  
 Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nilesanger.ac.uk  
 Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (4 kb). The v + 1 method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).  
 Email: nilesayed@tigr.org  
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T-brucei/.

FEATURES  
 source 1..26  
 /organism="Trypanosoma brucei"  
 /strain="TREU927"  
 /db\_xref="taxon:5691"  
 /clone="194f01"

BASE COUNT 0 a 3 c 11 g 12 t  
 ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 26;  
 Best Local Similarity 78.9%; Pred. No. 9.6e+04;  
 Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 acaccgcctctcgcaaa 23  
||||| | | | | | | | |  
Db 21 ACACCGCGCACACGACGACA 3

RESULT 13  
A2845779/c 35 bp DNA GSS 20-FEB-2001  
LOCUS 2M0145B13R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
DEFINITION clone UUGC2M0145B13 R, DNA sequence.  
ACCESSION A2845779  
VERSION A2845779.1 GI:13015687  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 35)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,  
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0145 row: B column: 13  
Seq primer: CACACGACGAAACACGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 35.  
FEATURES  
source Location/Qualifiers  
1..35  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0145B13"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PMD42 (g1473214|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

BASE COUNT 3 a 1 c 15 g 16 t  
ORIGIN

Query Match 54.8%; Score 12.6; DB 13; Length 35;  
Best Local Similarity 78.9%; Pred. No. 1e+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 acaccgcctctcgcaaa 23  
||||| | | | | | | | |  
Db 35 ACACCGCGCTCACACACACA 17

RESULT 14  
AUI06768/c 50 bp mRNA EST 05-APR-2001  
LOCUS AUI06768 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone  
DEFINITION HEP11938, mRNA sequence.  
ACCESSION AUI06768  
VERSION AUI06768.1 GI:13556289  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 50)  
AUTHORS Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hata  
'H., Ota,T., Isogai,T., Tanaka,T., Nakamura,T., Morishita,S., Okudo  
'K., Suyama,A. and Sugano,S.  
Fine structural analysis of transcription start sites of human  
mRNAs using full-length enriched and 5'-end enriched cDNA libraries  
unpublished (2001)  
JOURNAL Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yszuk@iems.u-tokyo.ac.jp  
Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano  
'S. Construction and characterization of a full length-enriched and  
a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).  
FEATURES  
source Location/Qualifiers  
1..50  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="HEP11938"  
/clone\_lib="Sugano Homo sapiens cDNA library"  
BASE COUNT 12 a 11 c 17 g 10 t  
ORIGIN

Query Match 54.8%; Score 12.6; DB 10; Length 50;  
Best Local Similarity 78.9%; Pred. No. 1e+05;  
Matches 15; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 agaaccgcctctcgcc 20  
||||| | | | | | | | |  
Db 40 AGCACCGCGCTCCCTCTC 22

RESULT 15  
A2462085 32 bp DNA GSS 04-OCT-2000  
LOCUS A2462085  
DEFINITION 1M0269P08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC1M0269P08 F, DNA sequence.  
ACCESSION A2462085  
VERSION A2462085.1 GI:10620210  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 32)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamll,C.,  
'M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss

University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
 84112, USA

Tel: 801 585 5606  
 Fax: 801 585 7177

Email: ddunne@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00

Plate: 0269 row: P column: 08  
 Seq primer: CGTGTAAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 32.

## FEATURES

## SOURCE

Location/Qualifiers

1..32

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUC1M0269P08"

/clone\_lib="Mouse 10Kb plasmid UUC1M library"

/sex="Male"

/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## BASE COUNT

7 a 14 c 1 g 10 t

## ORIGIN

## Query Match

Best Local Similarity 53.9%; Score 12.4; DB 13; Length 32;  
 Matches 13; Conservative 0; Mismatches 1; Indels 0;

QY 5 acaccgcctctc 18

DB 3 ACACCCACTCTCTC 16

Search completed: March 9, 2002, 00:09:25  
 Job time: 11041 sec